4-18-2023

An experimental look at reasonable suspicion and police discretion

Kyle McLean
Clemson University

Justin Nix
University of Nebraska at Omaha, jnix@unomaha.edu

Seth W. Stoughton
University of South Carolina

Ian T. Adams
University of South Carolina

Geoffrey P. Alpert
University of South Carolina - Columbia

Recommended Citation

McLean, K., Nix, J., Stoughton, S.W., Adams, I.T. and Alpert, G.P. (2023), "An experimental look at reasonable suspicion and police discretion", Policing: An International Journal, Vol. ahead-of-print No. ahead-of-print. https://doi.org/10.1108/PIJPSM-01-2023-0002

An experimental look at reasonable suspicion and police discretion

Forthcoming at *Policing: An International Journal**
doi: 10.1108/PIJPSM-01-2023-0002
(Accepted 3/17/2023)

Kyle McLean[1]
Justin Nix[2]
Seth W. Stoughton[3]
Ian T. Adams[4]
Geoffrey P. Alpert[4]

[1] Corresponding Author. College of Behavioral, Social, and Health Sciences, Clemson University. Email: kdmclea@clemson.edu.
[2] School of Criminology and Criminal Justice, University of Nebraska Omaha.
[3] School of Law, University of South Carolina.
[4] Department of Criminology and Criminal Justice, University of South Carolina.

emerald PUBLISHING

Policing: an International J

An Experimental Look at Reasonable Suspicion and Police Discretion

Journal:	*Policing: An International Journal*
Manuscript ID	PIJPSM-01-2023-0002.R1
Manuscript Type:	Research Paper
Keywords:	Policing, traffic stops, reasonable suspicion, Discretion

SCHOLARONE™
Manuscripts

ABSTRACT

Purpose: Demonstrate the need for further examination of legal judgments and the exercise of discretion in policing.

Design: A factorial vignette survey with traffic stop scenarios based on US Court of Appeals decisions was administered to 396 police officers across six states. Officers were asked to indicate their assessment of the presence of reasonable suspicion and the likelihood that they would extend the stop for investigatory purposes.

Findings: Officers' reasonable suspicion judgments are significantly influenced by the vignette facts and align with court ruling expectations. However, even in the presence of reasonable suspicion, responses indicate a limited use of officer discretion to extend the stop.

Originality: Analyses of officer decision-making often rely on large datasets with easy indicators of location, officer demographics, and citizen demographics, but rarely consider the facts of individual cases. Our study suggests more experimental research is needed to consider the impact of case facts on officer judgments and discretionary activity.

Keywords: policing, traffic stops, reasonable suspicion, police discretion

Policing is simultaneously governed by laws and subject to considerable discretion. While officers' authority to detain, search, arrest, and use force is regulated by constitutional standards, these standards overwhelmingly rely on the vague and flexible term "reasonable." For example, a temporary detention is allowed when an officer has *reasonable* suspicion that a crime has been committed (*Terry v. Ohio*; 392 U.S. 1, 1968). The use of force is constitutional when it is considered "objectively *reasonable*" (*Graham v. Connor*; 490 U.S. 386, 1989). Although much of the legal system relies on the concept of the *reasonable* person (Alpert & Smith, 1994; Fleming, 1951; Green, 1967), the specific nature of police enforcement actions—which can involve the deprivation of liberty, the invasion of privacy, and the use of force—makes reasonableness particularly salient.

Reasonableness, however, is inherently nebulous; often, the line between "reasonable" and "unreasonable" decisions can be indistinct. At its core, reasonableness is simply an assessment that the intrusion of a particular police action is balanced by the legitimate public safety interests of the government (Ashdown, 1981; Garrett, 2017; Stoughton et al., 2022). At a micro-level, when officers err in applying legal reasonableness, individuals who are *not* reasonably perceived as behaving suspiciously may be detained, and individuals whom officers may *un*reasonably believe committed a crime may be arrested. At a macro-level, judgments of reasonableness also factor into discussions of under- or over-policing (Atherley et al., 2022; Lewis & Usmani, 2022; Torres & Reling, 2020). Arguments regarding under-policing suggest that the police should be conducting more intrusions into individuals' lives based on a variety of contextual factors (e.g., crime rates), while over-policing suggests the opposite. The pervasiveness of these arguments in scholarly journals and public discourse is itself evidence that the line between what is reasonable and what is not is difficult to ascertain. And even when that

line is clear, police officers have considerable authority to exercise discretion in choosing whether to take enforcement actions (Brown, 1988). Not every individual who is "reasonably" perceived as behaving suspiciously must be detained, and not every individual whom officers "reasonably" believe committed a crime will be arrested.

Reasonableness is traditionally viewed through either a legal or sociological lens. The legal lens entrusts courts to determine reasonable actions, which are legally justified and permit enforcement actions. The sociological lens evaluates justifiability of stops based on citizen perceptions, outcome disparities, or data-driven criteria, often overlooking court-defined legal reasonableness standards. However, separating these lenses creates significant gaps in comprehension for both research areas. Legal scholars, by solely focusing on courts, neglect the societal context's impact on officers' reasonableness judgments and enforcement decisions. Conversely, social scientists, by concentrating only on social context and outcomes, disregard the fact that officers must meet legal reasonableness standards to engage in enforcement actions.

For example, whether due to data limitations, a lack of legal training, or nebulous criteria from the court, criminological studies of traffic stops (Alpert, Dunham, & Smith, 2007; Grogger & Ridgeway, 2006; Tillyer & Engel, 2013), pedestrian stops (Tillyer, Smith, & Lloyd, 2021), searches (Rojek, Rosenfeld, & Decker, 2012), and even stop and frisk (Gelman, Fagan, & Kiss, 2007) have all considered the legal actions that officers take using models that are unable to account for officers' judgments of legal reasonableness. These approaches have shed light on racial disparities and other issues in policing but have masked officers' consideration of the underlying factors that influence this decision-making process. These analyses can show that officers stop racial minorities at higher rates than whites. However, they cannot determine if this is the result of 1) officers' conscious discretionary decisions to take legal action against

individuals of color but not white individuals, 2) unconscious bias in the officers' perceptions that lead them to believe that individuals of color are *more* suspicious than similarly situated white individuals, or 3) a difference in how the officer operationalizes "reasonable" suspicion that leads them to use a lower threshold of suspicion for individuals of color than they do for white individuals. The latter findings (i.e., 2 or 3) would be consistent with a body of literature on implicit biases (Kahn & Martin, 2020; Smith & Alpert, 2007).

This study examines the limitations of analyzing officers' enforcement actions without considering their assessments of legal reasonableness, utilizing a factorial vignette survey design. The survey focuses on reasonable suspicion and discretionary enforcement activities related to potential drug activity in two vignette scenarios based on real cases from U.S. Courts of Appeals. In both instances, randomized fact manipulations significantly predict officers' assessments of reasonable suspicion and their willingness to engage in enforcement actions. Consequently, we advocate for policing scholars to employ diverse methods to investigate police discretion, taking into account legal factors that impact officers' judgments and behaviors.

Considering Reasonable Suspicion

Criminological studies of officers' legal actions have largely relied on big data sources (e.g., calls for service) to understand disparities in stops, searches, and arrests (Grogger & Ridgeway, 2006; Huff, 2021; Tillyer & Engel, 2013; Tillyer, Smith, & Lloyd, 2021), as well as the officer characteristics that influence the likelihood that officers will take enforcement action (Ba et al., 2021). Despite the useful results, these studies are limited by the use of administrative data. They, therefore, have not been able to consider how officers make legal judgments, including whether and why the enforcement action is properly predicated on probable cause or reasonable suspicion. Where research on the use of force has routinely grappled with the

influence of subject actions (e.g., threat, resistance, and demeanor) in influencing use of force outcomes (Alpert & Dunham, 2004; Alpert, Dunham, & MacDonald, 2004; Engel, Sobol, & Worden, 2000; Garner, Maxwell, & Heraux, 2002; Hine, Porter, Westera, & Alpert, 2018; McLean, Alikhan, & Alpert, 2022; Terrill, 2005), stop, search, and arrest studies commonly examine outcomes without considering this type of evidence (cf., Paoline & Terrill, 2007). Data limitations have often forced models of officers' enforcement activity to proceed without the inclusion of legal criteria, such as the evidence known to the officer. In short, these models ignore the concept of legal reasonableness in estimating disparities or the impact of various policies and training programs on enforcement actions.

Neglecting legal reasonableness restricts the influence of criminological research on the criminal justice system. For instance, in *Floyd v. City of New York* (959 F.Supp.2d 540, 2013), the case examined stop-question-and-frisk practices, utilizing statistical analyses similar to criminological studies, such as racial breakdowns of stops and frisk hit rates. Although low hit rates and significant racial disparities were concerning, the ultimate legal conclusion hinged on the judge's determination—backed by statistical analysis—that certain stops were unconstitutional due to insufficient reasonable suspicion. By considering both sociological context and legal reasonableness, this case significantly impacted stop-question-and-frisk practices in New York City. The judicial remedy did not prohibit all stops and frisks but required the agency to make changes to minimize constitutional violations through policy, supervision, monitoring, and training, as well as engaging in a joint remedy process with stakeholders. However, this combined sociological and legal analysis is more of an exception than the norm in such research.

While including considerations of legal reasonableness is critical for maximizing the impact of analyses of enforcement actions, part of the challenge of exploring these concepts in criminological research results not just from data deficiencies, but from a lack of clarity from the legal system. As noted by Stoughton and colleagues (2022), courts have traditionally and emphatically resisted bright-line rules that would make rigorous quantitative testing of officers' legal judgments methodologically sound. Courts require that reasonableness be based on *particularized* facts that are judged on a case-by-case basis (*Terry v. Ohio*, 392 U.S. 1, 21; 1968). As a result, there are scenarios where even having a law degree allows no more than a basis for contestation about whether borderline cases of suspicion are reasonable or not. Aside from this approach making it highly likely that officers—who, it must be said, typically do not have law degrees—will occasionally fail to recognize either the extent or the limits of their authority, it also precludes academics from being able to test officers' reasonableness judgments quantitatively, as these approaches typically require a large amount of data and cannot be judged case-by-case.

Even with these challenges, criminologists have long recognized the importance of officers' legal judgments of reasonableness and the exercise of police discretion (Brown, 1988; Davis, 1975). To this end, scholars have directed considerable attention towards "taming" police discretion and limiting enforcement activities to reduce abuses of power in police reform efforts (Walker, 1993). Ideas abound, such as limiting traffic stops for non-moving violations or the authority to conduct searches in certain circumstances (Epp, Maynard-Moody, & Haider-Markel, 2014; Fliss et al., 2020). In some ways, however, these approaches similarly sidestep the issue of reasonableness by creating bright-line rules for enforcement actions that are not based on the reasonableness of the determination that an offense has been committed, but rather upon policy

decisions. Admittedly, there are many factors that may influence judgments of reasonable suspicion and the exercise of discretion, however, we will focus on just three factors in the present study that are reasonably easy to manipulate experimentally: the evidence, suspect race, and suspect gender.

The Evidence

Legally, the primary driver of assessments of reasonableness and the exercise of discretion should be the facts of the case. While that may seem obvious, we have already noted that prior criminological work predominantly focuses on outcomes—assessing aggressiveness and disparities by examining overall rates of stops (Alpert et al., 2007; Grogger & Ridgeway, 2006; Tillyer & Engel, 2013; Tillyer, Smith, & Lloyd, 2021), searches (Rojek, Rosenfeld, & Decker, 2012), and arrests (Headley & Wright, 2020; Smith, Rojek, Petrocelli, & Withrow, 2017), rather than examining decisions in light of the facts of the case. So, for example, an analysis may consider the rates at which officers stop, ticket, and/or arrest individuals for driving under the influence, but the research is uninformed about whether the driver's behavior gave rise to suspicion that they *were*, in fact, under the influence. Yet, certainly it is more reasonable to suspect a driver of being impaired when they are failing to maintain a single lane, traveling exactly at the posted speed limit, and running stop signs than when they are simply driving at the posted speed limit. Indeed, studies in other areas of criminal justice have demonstrated that the strength and amount of evidence is related to outcomes such as jury verdicts (Devine et al., 2009) and plea bargains (Peterson et al., 2022).

With respect to policing, Knowles, Perisco, and Todd (2001) proposed a method for assessing bias in discretionary activities in policing through hit-rates for police searches—a rate of how often police searches resulted in contraband being found. However, Engel (2008) pointed

out that the test proposed should be used with caution as it neglected to distinguish between entirely discretionary searches (e.g., vehicle searches based on probable cause) and non-discretionary searches (e.g., searches of an individual when the person was already being arrested). These alternate views ultimately led to a rather lengthy debate (see Engel & Tillyer, 2008; Perisco & Todd, 2008) over the ability to conduct hit-rate tests for police searches. Regardless of a scholar's position on the issue, the debate only exists because of limitations of the available data. That is, clearly the facts before the officer matter in determining whether the search was both *legally* reasonable and whether it was socially appropriate or *normatively* reasonable, but debates over hit-rate tests occur because the facts before the officer are almost always unavailable to researchers assessing sociolegal processes in policing. Accordingly, we suggest that scholars should think more carefully about how officers consider the facts in front of them before making assessments of reasonable suspicion.

Suspect Race and Gender

Police reform efforts have long considered the importance of understanding suspects' characteristics—e.g., age, race, gender, or any other number of factors—in shaping officers' legal judgments. Prior studies have shown that racial minorities are more likely to be stopped (Ridgeway, 2007; Warren et al., 2006), searched (Pierson et al., 2020; Ridgeway, 2007), and have force (Kramer & Remster, 2018) used against them than whites. Women also experience differential treatment at the hands of police (Visher, 1983) with police typically taking a more legalistic (i.e., non-discretionary) approach to female suspects (Buzawa & Hotaling, 2006) and being less likely to use force (e.g., Garner et al., 2002). As noted previously, these influences may be the result of either conscious or unconscious biases in discretionary decisions or in the formation of legal judgments. For example, Warren and colleagues (2006) note the possibility

that disparities in stops could occur from racial profiling and prejudices (conscious biases), cognitive biases (subconscious biases), or differential police deployments in racial minority communities.

Officers' conscious or subconscious perceptions of minorities as more dangerous or suspicious may impact their assessments of whether legal thresholds for enforcement actions are met. For instance, recent quasi-experimental research has shown that traffic stops of black drivers increased near Trump rallies during the 2015-2016 presidential campaign. This effect was more pronounced in areas with historical slavery ties, more Jim Crow-era racial violence, and among officers with higher estimated baseline racial bias (Grosjean, Masera, & Yousaf, 2022). While these factors should not influence officers' legal judgments, scholars should investigate their potential effects to help policymakers and practitioners understand and mitigate their impact on legal decisions. In other words, while suspect characteristics should ideally be irrelevant to legal judgments and discretion, they must be considered to identify, reduce, or eliminate any influence they may have.

Current Study

This study aims to provide deeper insights into suspicion and discretion by employing a factorial vignette survey design with officers across six states. The survey specifically addresses reasonable suspicion and discretionary enforcement activity in two vignette scenarios based on real appellate court cases, focusing on potential drug activity. As an initial step, we examine how the amount of evidence presented, as well as the suspect's race and gender, influence officers' assessments of reasonable suspicion and their willingness to engage in enforcement activity. Importantly, this methodology allows us to advance by using standardized real-world scenarios

to estimate the causal impacts of (1) case facts on officers' legal judgments and enforcement action decisions, and (2) suspect race and gender on these outcomes.

Methods

Online surveys were administered to a convenience sample of 396 police officers from six departments located in diverse geographical areas with varying demographic backgrounds. The departments were not sampled from any meaningful sampling frame; rather, members of the research team contacted people whom they knew from these departments and asked for permission to administer the survey in each. These six agencies agreed to participate by sending an email solicitation to all of their sworn officers. In Spring 2022, 2,009 officers received the email solicitation, and 396 completed at least part of the survey, yielding a 19.7% response rate.

The voluntary nature of the survey, online administration mode, and lack of participation incentives likely explain why the response rate was not higher (Nix et al., 2019). However, randomization to vignette conditions occurred after officers agreed to participate and they had been informed about the survey's focus on suspicion formation and discretion. While randomization enables causal inference in treated groups, we cannot observe or determine if the experimental findings are affected by complete nonresponse—a common limitation of survey methodologies (for general discussion, see Groves & Peytcheva, 2011; Rindfuss et al., 2015).

Surveys were distributed via Qualtrics, with officers accessing the survey through a link in the solicitation email and proceeding to an informed consent form. After agreeing to participate, officers read a randomized vignette describing a traffic stop based on United States v. Pena-Gonzalez (618 F. App'x 195, 5th Cir. 2015). The vignette varied on three dimensions—amount of evidence, driver's race/ethnicity, and driver's gender—resulting in 18 possible conditions, with one randomly assigned using Qualtrics' random display function.

After reading the vignette, officers answered questions about their perception of the traffic stop, including their assessment of reasonable suspicion for extending the stop to investigate potential drug activity and their likelihood of doing so. Next, officers read a second randomized vignette based on United States v. Simpson (609 F.3d 1140, 10th Cir. 2010), which varied along the same dimensions as the first vignette and was randomly assigned separately. This second vignette was followed by the same questions. Finally, officers answered a series of demographic questions.

It is crucial to emphasize that the vignettes in this study are based on actual traffic stops that reached appellate courts. In both cases, the courts concluded that officers had reasonable suspicion to conduct and continue the stops, identifying specific pieces of evidence as necessary for that conclusion. The vignettes were designed so that officers would receive a vignette with either: less than the minimum amount of evidence the court found necessary to establish reasonable suspicion, the minimum amount of evidence the court deemed necessary for reasonable suspicion, or more than the minimum amount of evidence the court found necessary for reasonable suspicion.

Although these court decisions are somewhat debatable—particularly as they are circuit court opinions that may not apply in each agency's jurisdiction[1]—they still offer sufficient manipulations for expecting officers to make differential judgments on reasonable suspicion based on the evidence provided in the vignette. The exact wording of the vignettes can be found in the Supplemental Materials.

[1] For the scenario based on *United States v. Simpson*, one agency is located in the Tenth Circuit. For the scenario based on *United States v. Pena-Gonzalez* none of the participating agencies were located in the Fifth Circuit, and the opinion was "unpublished," meaning that it did not have precedential effect. Nevertheless, the two cases can be useful touchstones for courts even in other jurisdictions that have to address factually similar situations.

Measures

Dependent Variables

There are two primary dependent variables in this study. The first is the officer's *assessment of reasonable suspicion presence*, a dichotomous response (yes=1, no=0) to the question, "Do you have reasonable suspicion to believe that there are drugs either on the driver or in the vehicle, giving you discretionary authority to prolong this traffic stop to investigate possible drug activity?" The second dependent variable is the officer's *likelihood of extending the stop*, measured using a five-point scale in response to the question, "If you were the officer responding to this scenario, how likely or unlikely is it that you would extend the traffic stop to investigate possible drug activity?" The scale ranges from extremely unlikely (1) to extremely likely (5). For analysis purposes, the likelihood of extending the stop variable is collapsed into a dichotomous yes (1) or no (0) response. Responses of extremely unlikely (1), somewhat unlikely (2), and neither likely nor unlikely (3) are coded as no, while somewhat likely (4) and extremely likely (5) are coded as yes. Further justification for this decision is provided in the analytic strategy subsection.

Independent Variables

Evidence Presented.

As noted above, the scenarios were constructed consistent with federal court of appeals rulings to have either less than the evidence necessary to justify reasonable suspicion, exactly the minimum evidence needed to support reasonable suspicion, or more evidence than needed to support reasonable suspicion. These vignette conditions are broken down into a series of variables. Specifically, the evidence conditions are dummied into *less than minimal* evidence and

enhanced evidence, both referenced against the middle category of the exact right evidence supporting reasonable suspicion (as determined by court rulings discussed above).

Suspect Race and Gender

The vignette manipulations also randomly manipulated the *race/ethnicity* of the driver, with the driver being described as "white," "black," or "Hispanic," as well as the *gender* of the driver ("male" or "female"). Race/ethnicity is dummied into *black* and *Hispanic* driver conditions, with white drivers serving as the reference category. Finally, gender is named *female,* with male drivers serving as the reference category.

Control Variables

A number of control variables from the survey are included to describe the sample that participated in the study. Their descriptive statistics can be found in Table 1. First, the *department* the survey was distributed in was coded into five dummy variables, with one department serving as the reference. Second, the officer's self-identified gender was asked with response options of male, female, non-binary, other, or prefer not to say. None of the 326 officers providing a valid response selected a category other than male or female. Accordingly, the variable was dichotomized as male or any other category (presented as *officer male,* where 1=male).[2] We also measured officers' primary assignment, broken down into four categories: patrol, *specialized field unit*, *investigative*, or *administrative*. Patrol serves as the reference category with the other three options dichotomized. To ensure survey generalizability across all seven sources, officer rank was measured in four categories: non-supervisory officer, *front-line supervisor*, *mid-level supervisor*, and *command supervisor* (non-supervisory officer serves as the

[2] Note that, as is common in policing, 70.2% of the sample identified as male, leaving just 12.1% of the sample as female, and 17.7% of the sample declining to respond.

reference category in all analyses).[3] Officer's education was collected in five categories: high school diploma, some college (no degree), *associate's degree*, *bachelor's degree*, and *graduate degree*. Only 15 participants selected high school diploma, so to avoid small cell issues in analyses, this category was combined with some college (no degree) to form a "no degree" category. These four education categories were then treated as a series of dummy variables with "no degree" serving as the reference category. Prior *military experience* is measured as a simple yes or no to the question, "Have you ever served in the military?" Finally, years of experience was collected as an open-ended integer response to the question, "How long (in years) have you worked in law enforcement?" A discussion of how the sample matches to national estimates of police departments provided by the Law Enforcement Management and Administrative Statistics Survey (Goodison, 2022) is included in the Supplemental Materials.

[Insert Table 1 About Here]

Analytic Strategy

We start by showing the percentage of officers within each experimental condition who found reasonable suspicion to be present and indicated they would likely extend the stop. Following that, we present a logistic regression of the experimental manipulations on both the reasonable suspicion and extend the stop outcomes to test for significant differences in legal judgments and enforcement actions based on the facts and suspects' demographic characteristics. Given that the vignettes are randomly assigned, we can estimate the effects of these

[3] An "other" category was included for both assignment and rank with officers able to write-in text of their assignment and rank. All write-in values were re-coded according to the research team's best interpretation of their input (e.g., detective was recoded to investigative assignment, captain was re-coded to command supervisor).

manipulations without control variables, and have a reasonable degree of certainty in the internal validity of these causal explanations.[4]

Analyses of willingness to extend the stop were originally attempted as ordered logistic regressions, but parallel slopes assumptions were violated in both ordered logistic and generalized ordered logistic regressions. Accordingly, we shifted to a multinomial logistic regression approach. In testing these models, Wald tests for combining response options were conducted (Long & Freese, 2014), and results indicate no significant differences between option choices 1, 2, and 3, or 4 and 5. Thus, these options were combined as indicated above in the measures subsection and the analyses reverted to a logistic regression.

Results

[Insert Figure 1 About Here]

Most notable from the descriptive results of the vignette conditions in Figure 1 is a clear indication that, compared with Scenario 1, officers are more likely both to rate Scenario 2 as having reasonable suspicion present and to want to extend the stop in Scenario 2 across all vignette conditions.[5] Furthermore, the two bars within each scenario are nearly identical with little apparent variation between officers' judgment of reasonable suspicion and their likelihood of taking enforcement action. Accordingly, Table 2 further investigates this relationship and shows an incredibly strong relationship between officers' ratings of reasonable suspicion and their likelihood of extending the stop. In Scenario 1, only 17.6% of officers who indicated that reasonable suspicion was present indicated they were not likely to extend the stop—thereby

[4] A supplemental analysis that includes the control variables is included in the Supplemental Materials. However, the findings support the conclusions presented here.

[5] This difference is also statistically significant. Officers were less likely to indicate reasonable suspicion was present in Scenario 1 (M=0.30, SD=0.46) than in Scenario 2 (M=0.71, SD=0.45), $t(337)=-13.28$, $p<0.000$. Additionally, officers were less likely to indicate a willingness to extend the stop in Scenario 1 (M=0.32, SD=0.47) than in Scenario 2 (M=0.69, SD=0.46), $t(338)=-11.77$, $p<0.000$.

indicating a willingness to exercise discretionary authority—while 10.7% of officers who indicated that reasonable suspicion was not present indicated they were at least somewhat likely to extend the stop despite not having the legal justification for doing so. Similarly, in Scenario 2, only 10.4% of officers who believed reasonable suspicion was present reported they were unlikely to extend the stop. In comparison, 17.5% of officers who did not believe reasonable suspicion was present indicated being at least somewhat likely to extend the stop.[6]

[Insert Table 2 About Here]

The experimental manipulations in the vignette demonstrate a clear impact of the evidence on officer perceptions of reasonable suspicion and willingness to extend the stop, but only mixed evidence for suspect race's impact and no support for suspect gender's impact (see Table 3). Having less than the minimum amount of evidence than the court of appeals determined was necessary to justify reasonable suspicion reduces the likelihood of officers indicating reasonable suspicion was present in both scenarios (Scenario 1: $b = -1.67$, $p < 0.01$; Scenario 2: $b = -0.95$, $p < 0.01$). Similarly, this less than minimal evidence condition reduces the likelihood of officers indicating they would be likely to extend the stop in both scenarios (Scenario 1: $b = -1.08$, $p < 0.01$; Scenario 2: $b = -0.97$, $p < 0.01$). Interestingly, having more evidence than the court of appeals determined was necessary to justify reasonable suspicion did not increase the likelihood of officers indicating reasonable suspicion was present or their willingness to extend the stop in either scenario. Suspect race had mixed evidence with race having a significant impact on the assessment of reasonable suspicion in Scenario 1 (Black: $b = -0.67$, $p < 0.05$) and the decision to extend the stop in Scenario 2 (Black: $b = -0.73$, $p < 0.05$;

[6] Statistically, t-tests for differences of proportions indicates that the difference between the proportion of officers indicating a willingness to extend the stop is not significantly different from the proportion of officers indicating reasonable suspicion is present in either scenario. Scenario 1: $t(349)=1.04$, $p=0.30$. Scenario 2: $t(337)=-1.24$, $p=0.22$.

Hispanic: $b = -0.67$, $p < 0.05$). However, in each case, these findings are in the ***opposite*** direction of what was expected with officers less likely to determine reasonable suspicion is present and being willing to extend the stop for minorities. Finally, the suspect gender manipulation has no impact on officers' perceptions of reasonable suspicion nor their likelihood of extending the stop in Scenario 1. In addition to these limited findings, the pseudo r-squared values across all models are relatively small—all are less than 0.1.

[Insert Table 3 About Here]

Discussion

The constitutional standard governing the use of coercive authority in policing—covering the use of force, brief detentions, arrests, frisks, or searches—mandates that officers can only infringe on individuals' protected liberty and privacy interests when it is objectively reasonable. The ambiguity in defining reasonableness has led legal scholars to describe it as a "deeply impoverished" (Harmon, 2008, p. 119) "mess" (Dworkin, 1973, p. 329) of "embarrassing" (Amar, 1994, p. 785) rules constituting a "tangled skein" (Stoughton, 2014, p. 850) of constitutional jurisprudence. Simplifying the complexities, officers can undertake specific enforcement actions when the available facts establish "certain levels of suspicion," referred to as "reasonable suspicion" and "probable cause" (Stoughton et al., 2022, p. 39). In practice, this demands that officers make subjective judgments about the objective legal standard. However, policing researchers often struggle to account for the impact of case facts when assessing officers' discretionary enforcement activities.

This study demonstrates the necessity of considering individual case facts when examining officer decision-making. By experimentally manipulating evidence, suspect race, and suspect gender, we found that varying amounts of evidence influenced officers' judgments of

reasonable suspicion and their willingness to engage in enforcement action. However, additional evidence beyond the minimum required by the courts did not further increase these indications. While suspect race showed mixed results, all significant findings were contrary to prior literature's expectations. Suspect gender was unrelated to officers' assessments of reasonable suspicion and their likelihood of extending the stop.

The findings highlight the importance of the individual facts of a case on officer decision-making. Interestingly, our study reveals a relative lack of variance between reasonable suspicion and the likelihood of extending the stop. When officers determine that reasonable suspicion is present, they almost always indicate that they would likely extend the stop (Scenario 1=85.0%; Scenario 2=91.4%). This congruence suggests that, in the officer's mind, discretionary decisions are likely to align with the limits of their perceived legal authority. Furthermore, there seem to be more contextual factors influencing officer decision-making than those identified by the courts, as Scenario 2 had a higher percentage of officers indicating the presence of reasonable suspicion and a willingness to extend the stop than Scenario 1 across all vignette conditions.

This study has several limitations. First, the sample of participating officers came from a convenience sample of agencies willing to work with the authors, so the findings may not generalize to other agencies or officers who did not participate. Second, the study had a 19.7% response rate, likely due to the online administration method and lack of incentives offered to officers (Nix et al., 2019). Third, the study relied on written vignettes, which are a poor substitute for real-world experiences, and this artificiality might explain some of the findings, such as the discrepancy with previous research on race and enforcement activity.

The use of a survey vignette specifying race and gender may have primed officers to suppress or compensate for any biases, or be concerned about appearing biased, when indicating

the presence of reasonable suspicion or their likelihood of extending the stop. Using a between-subjects experimental design helps minimize this possibility, but social desirability bias may still mean that survey methods are an insufficient tool for accurately estimating race and gender-based implicit bias. Furthermore, this study only manipulated suspect race, suspect gender, and the evidence present in the encounter. A litany of other factors could drive officer discretion and decision-making, such as offender age, seriousness of the crime, presence of witnesses, etc. While the study's methodology maximizes our confidence in the internal validity of the results, it does not suggest that these other factors are unimportant. Instead, they represent a fruitful area for conducting additional experimental research in the future.

This study aims to encourage future research into legal judgments and discretion and should not be seen as the final word on how race and gender shape discretionary officer actions. The arguments and methodology presented here provide a solid foundation for re-evaluating the importance of evidence in making legal judgments and exercising discretion. While big data studies of large call for service databases have been useful, their dominance in the policing research landscape has led to a lack of consideration for the specific details of individual cases. This study highlights the importance of evidence and the facts of the case when officers make judgments of reasonableness, suggesting that big data studies might miss crucial factors in studying enforcement actions. Big data studies still hold value and should continue, but it is essential to explore additional research methods to gain a more comprehensive understanding of legal judgments and enforcement actions in policing. Criminologists should take advantage of the vague approach to reasonableness put forth by the court system and use empirical research to define what officers believe to be reasonable through rigorous research methods.

References

Alpert, G.P. & Dunham, R.G. (2004). *Understanding police use of force: Officers, suspects and reciprocity*. New York, NY: Cambridge University Press.

Alpert, G.P., Dunham, R.G., & Smith, M.R. (2007). Investigating racial profiling by the Miami-Dade Police Department: A multimethod approach. *Criminology & Public Policy, 6*, 25-55.

Alpert, G.P. & Smith, W.C. (1994). How reasonable is the reasonable man: Police and excessive force. *Journal of Criminal Law & Criminology, 85,* 481-501.

Amar, A.R (1994). Fourth Amendment first principles. *Harvard Law Review, 107*, 757-819.

Ashdown, G.G. (1981). The fourth amendment and the "legitimate expectation of privacy." *Vanderbilt Law Review, 34*, 1289-1345.

Atherly, L.T., Hickman, M.J., Parkin, W.S., & Helfgott, J.B. (2022). Measurement of potential over- and under-policing in communities. *Policing: A Journal of Policy and Practice, 16*, 443-461.

Auspurg, K., & Hinz, T. (2014). *Factorial Survey Experiments*. SAGE Publications.

Ba, B.A., Knox, D., Mummolo, J., & Rivera, R. (2021). The role of officer race and gender in police-civilian interactions in Chicago. *Science, 371*, 696-702.

Brown, M.K. (1988). *Working the street: Police discretion and the dilemmas of reform.* New York: Russell Sage Foundation.

Buzawa, E.S. & Hotaling, G.T. (2006). The impact of relationship status, gender, and minor status in the police response to domestic assaults. *Victims & Offenders, 1*, 323-360.

Davis, K.C. (1975). *Police discretion.* St. Paul: West Publishing Co.

Devine, D.J., Buddenbaum, J., Houp, S., Studebaker, N., & Stolle, D.P. (2009). Strength of evidence, extraevidentiary influence, and the liberation hypothesis: Data from the field. *Law & Human Behavior, 33*, 136-148.

Dworkin, R.B. (1973). Fact style adjudication and the Fourth Amendment: the limits of lawyering. *Indiana Law Journal, 48*, 329-368.

Engel, R.S. (2008). A critique of the "outcome test" in racial profiling research. *Justice Quarterly, 25*, 1-36.

Engel, R.S., Sobol, J.J., & Worden, R.E. (2000). Further exploration of the demeanor hypothesis: The interaction effects of suspects' characteristics and demeanor on police behavior. *Justice Quarterly, 17*, 235-258.

Engel, R.S. & Tillyer, R. (2008). Searching for equilibrium: The tenuous nature of the outcome test. *Justice Quarterly, 25*, 54-71.

Epp, C. R., Maynard-Moody, S., & Haider-Markel, D. P. (2014). *Pulled over: How police stops define race and citizenship*. University of Chicago Press.

Fleming, J. (1951). The qualities of the reasonable man in negligence cases. *Missouri Law Review, 16*, 1-26.

Fliss, M.D., Baumgartner, F., Delamater, P., Marshall, S., Poole, C., & Robinson, W. (2020). Re-prioritizing traffic stops to reduce motor vehicle crash outcomes and racial disparities. *Injury Epidemiology, 7*. DOI: 10.1186/s40621-019-0227-6.

Garner, J.H., Maxwell, C.D., & Heraux, C.G. (2002). Characteristics associated with the prevalence and severity of force used by the police. *Justice Quarterly, 19*, 705-746.

Garrett, B. L. (2017). Constitutional reasonableness. *Minnesota Law Review, 102*, 61-126.

Gelman, A., Fagan, J., & Kiss, A. (2007). An analysis of the New York City Police Department's "Stop-and-Frisk" policy in the context of claims of racial bias. *Journal of the American Statistical Association, 102*, 813-823.

Goodison, S.E. (2022). *Local police departments personnel, 2020*. Office of Justice Programs: Bureau of Justice Statistics. Retrieved from: https://bjs.ojp.gov/data-collection/law-enforcement-management-and-administrative-statistics-lemas.

Green, E. (1968). The reasonable man: Legal fiction or psychosocial reality. *Law and Society Review, 2*, 241-258.

Grogger, J. & Ridgeway, G. (2006). Testing for racial profiling in traffic stops from behind a veil of darkness. *Journal of the American Statistical Association, 101*, 878-887.

Grosjean, P., Masera, F., & Yousaf, H. (2022). Inflammatory Political Campaigns and Racial Bias in Policing. *The Quarterly Journal of Economics*, qjac037.

Groves, R. M., & Peytcheva, E. (2008). The impact of nonresponse rates on nonresponse bias: A meta-analysis. *Public Opinion Quarterly*, *72*(2), 167–189.

Harmon, R.A. (2008). When is police violence justified? *Northwestern University Law Review, 102*, 1119-1188.

Headley, A.M. & Wright, J.E. (2020). Is representation enough? Racial disparities in levels of force and arrests by police. *Public Administration Review, 80*, 1051-1062.

Hine, K.A., Porter, L.E., Westera, N.J., & Alpert, G.P. (2018). Too much or too little? Individual and situational predictors of police force relative to suspect resistance. *Policing and Society, 28*, 587-604.

Huff, J. (2021). Understanding police decisions to arrest: The impact of situational, officer, and neighborhood characteristics on police discretion. *Journal of Criminal Justice*, *75*, 101829.

Kahn, K.B., & Martin, K.D. (2020). The social psychology of racially biased policing: Evidence-based policy responses, *Policy Insights from the Behavioral and Brain Sciences, 7*, 107-114.

Knowles, J., Perisco, N., & Todd, P. (2001). Racial bias in motor vehicle searches: Theory and evidence. *Journal of Political Economy, 109*, 203-229.

Kramer, R. & Remster, B. (2018). Stop, frisk, and assault? Racial disparities in police use of force during investigatory stops. *Law & Society Review, 52*, 960-993.

Lewis, C. & Usmani, A. (2022). The injustice of under-policing in America. *American Journal of Law and Equality, 2*, 85-106.

Long, J.S. & Freese, J. (2014). *Regression models for categorical dependent variables using Stata* (3rd Ed.). College Station: Stata Press.

McLean, K., Alikhan, A., & Alpert, G.P. (2022). Re-examining the use of force continuum: Why resistance is not the only driver of use of force decisions. *Police Quarterly*. DOI: 10.1177/10986111211066353.

Nix, J., Pickett, J.T., Baek, H., & Alpert, G.P. (2019). Police research, officer surveys, and response rates. *Policing and Society, 29*, 530-550.

Paoline, E.A. & Terrill, W. (2007). Police education, experience, and the use of force. *Criminal Justice and Behavior, 34*, 179-196.

Perisco, N. & Todd, P.E. (2008). The hit rates test for racial bias in motor-vehicle searches. *Justice Quarterly, 25*, 37-53.

Petersen, K., Redlich, A.D., & Wilson, D.B. (2022). Where is the evidence? Comparing the effects of evidence strength and demographic characteristics on plea discounts. *Journal of Quantitative Criminology*. DOI: 10.1007/x10940-022-09555-8.

Pierson, E. Simouiu, C., Overgoor, J., Corbett-Davies, S., Jenson, D., Shoemaker, A., Ramachandran, V., Barghouty, P., Phillips, C., Shroff, R., & Goel, S. (2020). A large-

scale analysis of racial disparities in police stops across the United States. *Nature Human Behavior, 4*, 736-745.

Ridgeway, G. (2007). *Analysis of racial disparities in the New York Police Department's stop, question, and frisk practices.* Santa Monica, CA: Rand Corporation. Retrieved from: https://www.rand.org/content/dam/rand/pubs/technical_reports/2007/RAND_TR534.pdf.

Rindfuss, R. R., Choe, M. K., Tsuya, N. O., Bumpass, L. L., & Tamaki, E. (2015). Do low survey response rates bias results? Evidence from Japan. *Demographic Research*, *32*, 797–828.

Rojek, J., Rosenfeld, R., & Decker, S. (2012). Policing race: The racial stratification of searches in police traffic stops. *Criminology, 50*, 993-1024.

Smith, M.R. & Alpert, G.P. (2007). Explaining police bias: A theory of social conditioning and illusory correlation. *Criminal Justice and Behavior, 34*, 1262-1283.

Smith, M.R., Rojek, J.J., Petrocelli, M., & Withrow, B. (2017). Measuring disparities in police activities: A state of the review. *Policing: An International Journal, 40*, 166-183.

Stoughton, S.W. (2014). Policing facts. *Tulane Law Review, 88*, 847-898.

Stoughton, S.W., McLean, K., Nix, J., & Alpert, G. (2022). Policing suspicion: Qualified immunity and "clearly established" standards of proof. *Journal of Criminal Law & Criminology, 112*, 37-78.

Terrill, W. (2005). Police use of force: A transactional approach. *Justice Quarterly, 22*, 107-138.

Tillyer, R. & Engel, R.S. (2013). The impact of drivers' race, gender, and age during traffic stops: Assessing interactions terms and the social conditioning model. *Crime & Delinquency, 59*, 369-395.

Tillyer, R., Smith, M.R., & Lloyd, C.D. (2021). May I talk to you for a minute? An assessment of street stops and officer decision-making. *Crime & Delinquency.* DOI: 10.1177/00111287211047541.

Torres, J., & Relling, T. (2020). Under-policing and apprehensiveness toward stopping minorities across white and nonwhite officer post-Ferguson. *Journal of Ethnicity in Criminal Justice, 4*, 282-314.

Visher, C.A. (1983). Gender, police arrest decisions, and notions of chivalry. *Criminology, 21*, 5-28.

Walker, S. (1993). *Taming the system: The control of discretion in criminal justice, 1950-1990.* Oxford University Press on Demand.

Warren, P., Tomaskovic-Devey, D., Smith, W., Zingraff, M., & Mason, M. (2006). Driving while black: Bias processes and racial disparities in police stops. *Criminology, 44*, 709-738.

Court Cases

Floyd v. City of New York; 959 F.Supp.2d 540, 2013
Graham v. Connor; 490 U.S. 386, 1989
Terry v. Ohio; 392 U.S. 1, 1968
United States v. Pena-Gonzalez; 618 F. App'x 195, 5th Cir. 2015
United States v. Simpson; 609 F.3d 1140, 10th Cir. 2010

Table 1. Descriptive Statistics

	N	%
Department		
1	82	20.7
2	67	16.9
3	40	10.1
4	71	17.9
5	80	20.2
6	56	14.1
Gender		
Male	278	70.2
Female	48	12.1
Assignment		
Patrol	162	40.9
Specialized Field Unit	49	12.4
Investigative	81	20.5
Administrative	35	8.8
Rank		
Non-Supervisory	219	55.3
Front-Line Supervisor	71	17.9
Mid-Level or Command Sup.	37	9.3
Education		
No degree	91	23.0
Associate's Degree	54	13.6
Bachelor's Degree	147	37.1
Graduate Degree	42	10.6
Military Experience		
Yes	87	22.0
No	246	62.1
Experience	Mean: 14.09	SD: 9.09

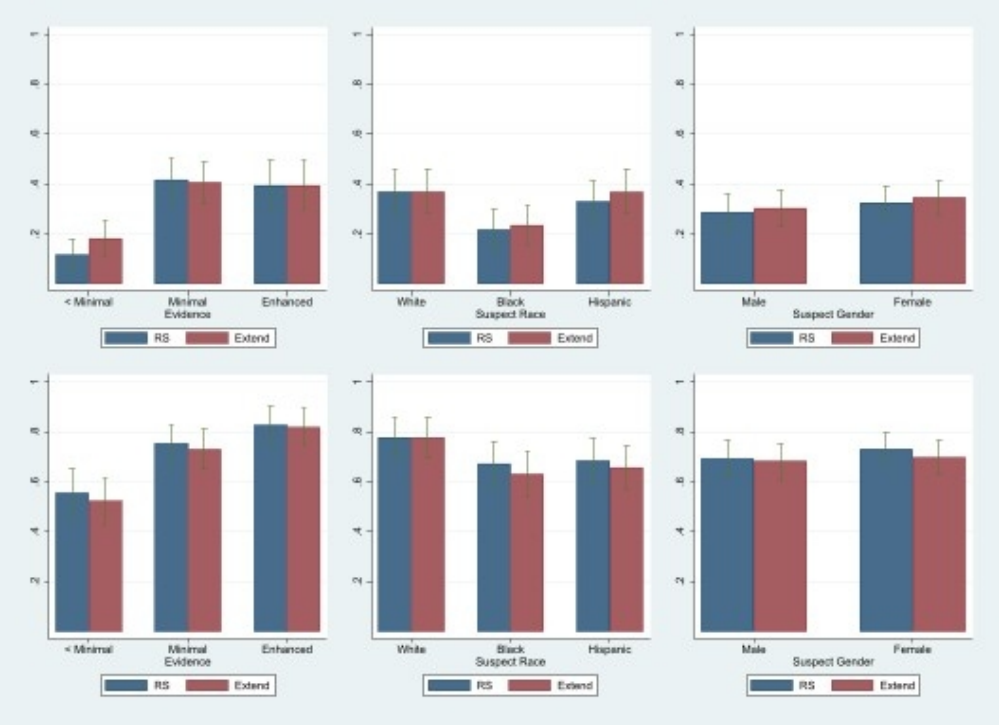
Table 2. Extending the Stop and Reasonable Suspicion

		Reasonable Suspicion	
		No	Yes
Extend the Stop	*Scenario 1*		
	No	216 (89.3%)	19 (17.6%)
	Yes	26 (10.7%)	89 (82.4%)
	Scenario 2		
	No	80 (82.5%)	25 (10.4%)
	Yes	17 (17.5%)	216 (89.6%)

Table 3. Experimental Results—Logistic Regression

	Reasonable Suspicion				*Extend the Stop*			
	Scenario 1		*Scenario 2*		*Scenario 1*		*Scenario 2*	
	b	S.E.	b	S.E.	b	S.E.	b	S.E.
Evidence								
Minimal	-1.67**	0.34	-0.95**	0.29	-1.08**	0.30	-0.97**	0.28
Enhanced	-0.10	0.27	0.38	0.33	-0.06	0.27	0.43	0.33
Driver Race								
Black	-0.67*	0.31	-0.54	0.31	-0.57	0.30	-0.73*	0.31
Hispanic	-0.29	0.29	-0.54	0.31	-0.08	0.27	-0.67*	0.31
Driver Gender								
Female	-0.13	0.25	0.18	0.25	0.14	0.24	0.05	0.25
N	350		338		351		339	
Chi-Squared	40.19**		24.88**		22.98**		29.82**	
Pseudo R-Squared	0.09		0.06		0.05		0.07	

*p<0.05, **p<0.01



Proportion of Officers Indicating Reasonable Suspicion/Willingness to Extend the Stop by Vignette Condition

139x101mm (96 x 96 DPI)

Supplemental Materials

Vignette Manipulations

Scenario 1, Enhanced Condition

While on patrol, you observe a vehicle driving two miles an hour over the posted limit on a highway frequently used by drug traffickers. There are four air fresheners hanging throughout the vehicle, three rosaries hanging from the rearview mirror, and four bumper stickers indicating support for D.A.R.E. and law enforcement. The vehicle is driving two miles an hour over the posted limit on a highway frequently used by drug traffickers. You stop the vehicle and question the driver, a white/black/Hispanic male/female. The driver gives inconsistent information about where he/she lives and his/her activities.

Scenario 1, Minimal Condition

While on patrol, you observe a vehicle driving slightly over the posted limit on a highway frequently used by drug traffickers. There are four air fresheners hanging throughout the vehicle. You stop the vehicle and question the driver, a white/black/Hispanic male/female. The driver gives inconsistent information about where he/she lives and his/her activities.

Scenario 1, Less than Minimal Condition

While on patrol, you observe a vehicle driving slightly over the posted limit on a highway frequently used by drug traffickers. There are four air fresheners hanging throughout the vehicle. You stop the vehicle and question the driver, a white/black/Hispanic male/female.

Scenario 2, Enhanced Condition

While patrolling a highway commonly used for transporting drugs, you see a vehicle driving well below the speed limit and with its windows rolled down despite it being a very hot day. The vehicle changes lanes twice without properly signaling, and you pull it over. When you approach, you see a butane lighter in the ashtray and butane lighter refills in the rear pocket of the front seat. There is also a radar detector pushed partly under the seat. When the driver, a white/black/Hispanic male/female, opens his/her glove compartment to retrieve his/her registration and insurance, you see a package of energy pills. The driver seems extremely nervous; his/her whole body is shaking. You say that you are only planning on issuing a ticket, but the driver remains nervous. When you ask about the driver's travel plans, he/she is evasive and provides inconsistent information. You learn from dispatch that the driver has previously been charged in another state for transporting drugs. You ask the driver whether they will answer more questions, but the driver says no.

Scenario 2, Minimal Condition

While patrolling a highway commonly used for transporting drugs, you see a vehicle change lanes twice without properly signaling, and you pull it over. The driver, a white/black/Hispanic male/female, seems extremely nervous; his/her whole body is shaking. You say that you are only planning on issuing a ticket, but the driver remains nervous. When you ask about the driver's travel plans, he/she is evasive and provides inconsistent information. You learn from dispatch that the driver has previously been charged in another state for transporting drugs.

Scenario 2, Les than Minimal Condition

While patrolling a highway commonly used for transporting drugs, you see a vehicle change lanes twice without properly signaling, and you pull it over. The driver, a white/black/Hispanic

male/female, seems extremely nervous; his/her whole body is shaking. You say that you are only planning on issuing a ticket, but the driver remains nervous. When you ask about the driver's travel plans, he/she is evasive and provides inconsistent information.

Sample Demographics Compared to LEMAS

With respect to the representativeness of the sample, 12% of our sample were female, which approximately matches the national average on female representation in law enforcement of 14% (Goodison, 2022). Just 41% of our sample were patrol officers, which is lower than the usual portion of officers in a local police department that work in patrol of approximately 65% (Goodison, 2022). At the same time, the percentage of officers working administrative assignments fell within the expected range (our sample: 8.8%; national averages: 4.7% to 10.7% depending on the size of the department, Goodison, 2022), but our sample had more investigators than expected (our sample: 20.5%; national averages: 14.7% to 11.1% depending on the size of the agency). Accordingly, our best assessment of the sample is that it falls in line with expectations from national surveys of policing, with the exception that investigators are over-represented.

Supplemental Analysis

To ensure accuracy of the findings presented in the manuscript, we have included the below supplemental analysis that includes control variables in the assessment of reasonable suspicion and the decision to extend the stop. These control variables include all of the descriptive statistics presented in Table 1 of the main manuscript, as well as five attitudinal scales. These scales are as follows:

Order Maintenance Attitudes. Consistent with prior work on police culture, o*rder maintenance* is a mean-scale score of six items asking officers, "How often do you think that patrol officers should be expected to do something about…" (1) neighbor disputes, (2) family disputes, (3) public nuisances, (4) nuisance businesses, (5) parents who do not control their kids, and (6) litter and trash from never (1) to always (5; α=0.78).

Law Enforcement Attitudes. Consistent with prior work on police culture, *law enforcement* is a single-item measure asking officers their level of agreement with the statement, "Enforcing the law is by far a patrol officer's most important responsibility."

Aggressive Patrol Attitudes. Similarly, *aggressive patrol* is also a single-item measure asking officers their level of agreement with the statement, "A good patrol officer is one who patrols aggressively by stopping cars, checking out people, running license checks, and so forth." Responses to both of the previous statements range from strongly disagree (1) to strongly agree (5).

Selective Enforcement Attitudes. Finally, *selective enforcement* is a single-item indicator asking officers, "How frequently would you say there are good reasons for not arresting someone who has committed a minor criminal offense?" using the same response options as the order maintenance scale. Critically, *selective enforcement* is the only item from this section modified

from prior work. This was done to emphasize discretionary decisions that are the focus of this research—i.e., having the legal authority to do something but using their discretion to not take the action.

Agency Aggressiveness. Officers were asked a series of questions about their agency's approach to strictly enforcing the law or exercising discretion to avoid making arrests. *Agency aggressiveness* is a mean-scale score of four of these items[1] such that higher scores indicate stricter enforcement of the law (α=0.62).[2]

Results from the inclusion of control variables on reasonable suspicion assessments can be seen in Table S-1. Critically, none of the effects of the experimental manipulations are substantively different (significance and direction remain unchanged and effect sizes are only slightly altered). Among the control variables, only one agency demonstrated a significantly lower likelihood of indicating reasonable suspicion was present. Furthermore, officers with military experience were somewhat more likely to indicate reasonable suspicion was present in the first scenario.

Table S-1. Logistic Regression on Presence of Reasonable Suspicion with Controls

	Scenario 1		*Scenario 2*	
	b	S.E.	b	S.E.
Evidence				
Minimal	-2.06**	0.40	-1.34**	0.35
Enhanced	-0.35	0.33	0.31	0.38
Driver Race				
Black	-0.85*	0.38	-0.47	0.36
Hispanic	-0.47	0.34	-0.55	0.36
Driver Gender				
Female	-0.22	0.30	0.21	0.29

[1] Items in this scale include: "My agency encourages officers to make arrests whenever possible" and "My agency encourages officers to strictly enforce all drug laws" coded from 1 (strongly disagree) to 5 (strongly agree), as well as "My agency encourages officers to find alternatives to arrest whenever possible" and "My agency encourages officers to use discretion regarding minor drug infractions" which are reverse-coded on the same scale.

[2] While these items fail to meet older thresholds for adequate internal consistency, modern scholars have suggested that these "alpha estimates make scales appear much less reliable than they are in actuality," (McNeish, 2018, p. 423). Further, these estimates are inflated when drawing respondents from different populations (i.e., different agencies), and therefore researchers should not hew too closely to traditional benchmarks of alpha reliability (Waller, 2008; Dunn, Baguley, & Brunsden, 2014).

Order Maintenance	0.27	0.23	0.05	0.22
Law Enforcement	-0.07	0.16	-0.01	0.15
Aggressive Patrol	0.31	0.17	0.20	0.16
Selective Enforcement	0.21	0.20	0.01	0.20
Agency Aggressiveness	0.16	0.24	-0.01	0.24
Source				
2	0.42	0.49	-0.39	0.52
3	-0.04	0.53	-0.78	0.54
4	0.10	0.49	-0.15	0.50
5	-1.50*	0.65	-1.40*	0.57
6	0.84	0.50	0.55	0.55
Officer Male	0.41	0.47	-0.45	0.45
Assignment				
Specialized Field Unit	0.39	0.43	0.07	0.42
Investigative	0.20	0.39	0.51	0.39
Administrative	0.32	0.60	0.79	0.62
Rank				
Front-Line Supervisor	0.55	0.38	0.42	0.38
Mid-Level/Command Supervisor	-0.71	0.60	-0.08	0.59
Education				
Associate's Degree	0.39	0.46	-0.38	0.45
Bachelor's Degree	0.21	0.37	-0.04	0.37
Graduate Degree	0.62	0.55	-0.44	0.54
Military Experience	0.67*	0.33	0.54	0.36
Experience	-0.01	0.02	-0.01	0.02
N	311		310	
Chi-Squared	79.71**		57.20**	
Pseudo R-Squared	0.21		0.15	

Table S-2 presents the results of the inclusion of control variables on the analysis of willingness to extend the stop. As before, the impacts of the experimental variables remain largely unchanged. The notable exception is that suspect race is reduced to non-significance in the second scenario. However, the effect size estimates are similar (Scenario 1: b = -0.73 without controls and b = -0.42 with controls; Scenario 2: b = -0.67 without controls, b = -0.47 with controls). With respect to the added control variables, officers' aggressive patrol attitudes are associated with a higher likelihood of being willing to extend the stop (Scenario 1: b = 0.39,

Scenario 2: $b = 0.41$). Again, one agency (the same as before) had a reduced likelihood of being willing to extend the stop, though it was joined by a second agency with an increased likelihood of being willing to extend the stop in the first scenario. Finally, male officers and more experienced officers were more likely to indicate a willingness to extend the stop in the first scenario.

Table S-2. Logistic Regression on Willingness to Extend the Stop with Controls

	Scenario 1		*Scenario 2*	
	b	S.E.	b	S.E.
Evidence				
Minimal	-1.78**	0.37	-1.31**	0.35
Enhanced	-0.23	0.34	0.44	0.39
Driver Race				
Black	-0.48	0.37	-0.42	0.36
Hispanic	-0.18	0.35	-0.47	0.36
Driver Gender				
Female	0.34	0.30	-0.11	0.29
Order Maintenance	0.32	0.22	0.25	0.22
Law Enforcement	-0.20	0.16	-0.15	0.15
Aggressive Patrol	0.39*	0.17	0.41**	0.17
Selective Enforcement	0.04	0.20	0.05	0.20
Agency Aggressiveness	-0.14	0.24	-0.41	0.24
Source				
2	0.35	0.49	0.48	0.53
3	0.24	0.52	-0.63	0.54
4	0.21	0.48	-0.13	0.48
5	-2.25**	0.69	-1.99**	0.57
6	0.99*	0.51	0.98	0.58
Officer Male	1.11*	0.51	-0.23	0.44
Assignment				
Specialized Field Unit	0.37	0.42	-0.14	0.43
Investigative	-0.26	0.39	0.33	0.40
Administrative	0.24	0.60	0.11	0.61
Rank				
Front-Line Supervisor	-0.18	0.39	0.66	0.39
Mid-Level/Command Supervisor	-0.64	0.59	0.27	0.62
Education				
Associate's Degree	0.32	0.46	-0.17	0.46
Bachelor's Degree	0.18	0.37	-0.06	0.37
Graduate Degree	0.14	0.56	-0.36	0.54

Military Experience	0.51	0.33	-0.43	0.35
Experience	0.04*	0.02	-0.02	0.02
N	311		311	
Chi-Squared	85.00**		68.38**	
Pseudo R-Squared	0.22		0.18	